

Entropy and Hausdorff Dimension in Random Growing Trees

Anna Rudas

Institute of Mathematics, Technical University of Budapest

rudasa@math.bme.hu

and

Imre Péter Tóth

MTA-BME Stochastics Research Group

and Department of Mathematics and Statistics, University of Helsinki

mogy@math.bme.hu

June 9, 2011

Abstract

We investigate the limiting behaviour of random tree growth in preferential attachment models. The tree stems from a root, and we add vertices to the system one-by-one at random, according to a rule which depends on the degree distribution of the already existing tree. The so-called *weight function*, in terms of which the rule of attachment is formulated, is such that each vertex in the tree can have at most K children.

We define the concept of a certain random measure μ on the leaves of the limiting tree, which captures a global property of the tree growth in a natural way. We prove that the Hausdorff and the packing dimension of this limiting measure is equal and constant with probability one. Moreover, the local dimension of μ equals the Hausdorff dimension at μ -almost every point. We give an explicit formula for the dimension, given the rule of attachment.

1 Introduction

We investigate a family of tree growth models in which the tree stems from a root in the beginning, and vertices are added one at a time, the new vertex always attaching to exactly one already existing vertex. The rule by which the new vertex chooses its parent, is dependent on the degree distribution apparent in the tree at the time the vertex is born. The models can be either in discrete time, when a vertex is born in every second, or in continuous time, then birth times are random. For the problems we discuss, these two versions are equivalent and can be translated into each other (details in Section 2.1). The classical models and results of the area use the discrete setting. However, for the proofs we give, the continuous-time version is much more natural and convenient, so this is what we will use.

This big family of models includes the Barabási-Albert graph [1] for example, in which the linear preferential attachment rule reproduces certain phenomena observed in real-world networks (e.g. the power law decay of the degree sequence). This property of the Barabási-Albert graph was proved in a mathematically precise way in [5] and, independently, in [18]. A wider class of models is considered in [15, 14], for rigorous results on different cases of this model, see [19, 21].

The results mentioned above focus on the local behaviour of the random tree, namely, they give results concerning the neighbourhood of a uniformly random vertex, which is chosen from the tree after a long time of tree evolution. In this paper we concentrate on global properties of the limiting tree.

It is natural to pose the following question. Let us fix a vertex, say the first vertex in the first generation, just above the root. What is the “limiting success level” of this vertex, compared to the other vertices in the same generation? What we mean by this is the number of descendants of this vertex, after a long time of tree evolution, compared to the number of descendants of its brothers.

Another formulation of the same question is to fix a vertex, let the tree grow for a long time, then choose a vertex uniformly at random from the big tree, and ask the probability that this random vertex is descendant of the fixed vertex. Clearly, if we look at these limiting probabilities for let us say the first generation, we get a distribution, itself being random, that codes an important information of the evolution of the tree.

If one looks at the system of these limiting (as time evolution of the tree tends to infinity) random distributions on the different generations of the tree, it is tempting to ask something about the limiting measure of this system, when letting the generation level tend to infinity. We will define the above concepts properly, and will denote this overall limiting measure by μ .

Having a random measure in our hand, which describes a global property of the limiting infinite system, it is natural to ask about the Hausdorff (and packing) dimension of this measure, for several reasons. First, these are the primary quantities capturing the scaling behaviour of the system, so they appear in statistical and Statistical Physics discussions. Secondly, these can actually be measured in (finite, but big) real systems, so they can be used to check the validity of models, or to tune parameters.

On the other hand, the dimension of the measure depends on a parameter of the underlying metric, which is arbitrary. To rule out this (trivial) dependence, it is usual to ask about the entropy of the limiting measure, which depends on the growth process only. This is the natural equivalent of the dimension from a dynamical point of view.

We prove the following results.

1. The limiting entropies (as time tends to infinity) of the random measures on the different generations converge to a constant with probability one, as we let the generation level to infinity. This constant h is called the entropy of the limiting measure μ .
2. The Hausdorff and the packing dimension of the random limiting measure μ are constant and equal with probability one. The entropy and the dimension satisfy the usual simple relation $dimension = \frac{entropy}{Ljapunov\ exponent}$, see (12). Moreover, the local dimension of μ equals the Hausdorff dimension at μ -almost every point.
3. Given the so-called *weight function* w , which determines the rule of the tree growth, we provide an explicit formula for the entropy, and thus for the Hausdorff dimension, in terms of w .

The key to these results is a Markov process appearing naturally in the construction of a μ -typical leaf of the tree. After some discussion of the tree structure, the Markov property will be easy to see. Some technical difficulties will arise from the non-compactness of the state space.

Our model is special in the sense that we only allow a finite degree for each vertex, but it is general in the sense that after having fixed the maximum number of children K a vertex may have, the weight function w , which determines the rule of attachment, can be any positive-valued function on $\{0, 1, \dots, K-1\}$.

The paper is structured as follows: The model and the results are presented in Section 2. This also includes a brief discussion of related models and related results in Section 2.5. Section 3 contains the main line of the argument, and ends with the proof of the first two results. Section 4 is devoted to proofs of lemmas which have been used but not proven in Section 3. Finally, Section 5 contains the proof of the last result.

2 Notation, Definitions and Results

We consider rooted ordered trees, which are also called family trees or rooted planar trees in the literature.

In order to refer to these trees it is convenient to use genealogical phrasing. The tree is thus regarded as the coding of the evolution of a population stemming from one individual (the root of the tree), whose “children” form the “first generation” (these are the vertices connected directly to the

root). In general, the edges of the tree represent parent-child relations, the parent always being the one closer to the root. The birth order between brothers is also taken into account, this is represented by the tree being an ordered tree (planar tree).

We only consider the case when every vertex can have at most $K \in \mathbb{N}$ children. We assume $K \geq 2$ to avoid the trivial case when only one child is born per parent. (In that case the tree growth is linear and the tree has no interesting structure.) We use the index set $\mathbb{I} := \{1, 2, \dots, K\}$, and also use $\mathbb{I}^- := \{0, 1, \dots, K-1\}$.

The vertices are labelled by the set

$$\mathcal{N} = \bigcup_{n=0}^{\infty} \mathbb{I}^n, \quad \text{where } \mathbb{I}^0 := \{\emptyset\},$$

as follows. \emptyset denotes the root of the tree, its first-born child is labelled by 1, the second one by 2, etc., and its last one by K , all the vertices in the first generation are thus labelled with the elements of \mathbb{I} . Similarly, in general, the children of $x = (i_1, i_2, \dots, i_n)$ are labelled by $(i_1, i_2, \dots, i_n, 1)$, $(i_1, i_2, \dots, i_n, 2)$, etc. Thus, if a vertex has label $x = (i_1, i_2, \dots, i_n) \in \mathcal{N}$, then it is the i_n^{th} child of its parent, which is the i_{n-1}^{th} child of its own parent and so on. If $x = (i_1, i_2, \dots, i_n)$ and $y = (j_1, j_2, \dots, j_l)$ then we will use the shorthand notation xy for the concatenation $(i_1, i_2, \dots, i_n, j_1, j_2, \dots, j_l)$, and with a slight abuse of notation for $i \in \mathbb{I}$, we use xi for $(i_1, i_2, \dots, i_n, i)$.

There is a natural partial ordering \prec on \mathcal{N} , namely, $x \prec z$ if x is ancestor of z , so if $\exists y \in \mathcal{N}$, $y \neq \emptyset$ such that $z = xy$. We use $x \preceq z$ meaning $x \prec z$ or $x = z$.

We can identify a rooted ordered tree with the set of labels of the vertices, since this set already identifies the set of edges in the tree. It is clear that a subset $G \subset \mathcal{N}$ may represent a rooted ordered tree iff $\emptyset \in G$, and for each $(i_1, i_2, \dots, i_n) \in G$ we have $(i_1, i_2, \dots, i_n - 1) \in G$ if $i_n > 1$, and $(i_1, i_2, \dots, i_{n-1}) \in G$ if $i_n = 1$.

We also think of \mathcal{N} as the complete rooted ordered tree.

\mathcal{G} will denote the set of all finite, rooted ordered trees. The *degree* of vertex $x \in G$ will denote the number of its children in G :

$$\deg(x, G) := \max\{i \in \mathbb{I} : xi \in G\} \quad (\text{zero if } x1 \notin G)$$

The *subtree* rooted at a vertex $x \in G$ is:

$$G_{\downarrow x} := \{y : xy \in G\},$$

this is just the progeny of x viewed as a rooted ordered tree.

2.1 The Model

2.1.1 Continuous-time Model

Given a function $w : \mathbb{I}^- \rightarrow \mathbb{R}_+$, referred to as the weight function, our randomly growing tree $\Upsilon(t)$ is a continuous-time, time-homogeneous Markov chain on the countable state space \mathcal{G} , with initial state $\Upsilon(0) = \{\emptyset\}$ and right-continuous trajectories.

The jump rates are the following. Suppose that at some $t \geq 0$ we have $\Upsilon(t-) = G$, then for each $x \in G$ which has $\deg(x, G) = j < K$, the process may jump to $G \cup \{xi\}$ with rate $w(\deg(x, G))$ where $i = j + 1$. This means that each existing vertex $x \in \Upsilon(t-)$ ‘gives birth to a child’ with rate $w(\deg(x, \Upsilon(t-)))$, independently of the others, and stops reproducing when reaches $\deg(x, \Upsilon(t)) = K$.

The Markov chain $\Upsilon(t)$ is well defined for $t \in [0, \infty)$, it does not blow up in finite time (see comment at (3)).

We define the *total weight* of a tree $G \in \mathcal{G}$ as

$$W(G) := \sum_{x \in G} w(\deg(x, G)).$$

Described in other words, the Markov chain $\Upsilon(t)$ evolves as follows: assuming $\Upsilon(t-) = G$, at time t a new vertex is added to it with total rate $W(G)$, and it is attached with an edge to exactly one

already existing vertex, which is $x \in G$ with probability

$$\frac{w(\deg(x, G))}{\sum_{y \in G} w(\deg(y, G))} .$$

2.1.2 Discrete-time Model

This continuous-time model naturally contains another, discrete-time model as follows. Define the stopping times

$$S_n := \inf\{t : |\Upsilon(t)| = n + 1\},$$

then the Markov chain $\Upsilon(S_n)$ is a randomly growing tree, where exactly one vertex is born at each time unit, and every newly born vertex chooses its parent at random, choosing x with probability

$$\frac{w(\deg(x, G))}{\sum_{y \in G} w(\deg(y, G))} .$$

if the $\Upsilon(S_{n-1}) = G$.

It was in this framework that Barabási and Albert originally formulated their model [1]. The relation of the two models is discussed in detail in [21]. As mentioned before, the questions we pose can be formulated equivalently in both models, but we will use the continuous-time version in our proofs, for reasons of convenience.

2.2 Some Additional Notation and Known Results

Let τ_x be the birth time of vertex x ,

$$\tau_x := \inf\{t > 0 : x \in \Upsilon(t)\} . \quad (1)$$

Let σ_x be the time we have to wait for the appearance of vertex x , starting from the moment that its birth is actually possible (e.g. when no other vertex is obliged to be born before him). Namely, let

- (a) $\sigma_\emptyset := 0$,
- (b) $\sigma_{y1} := \tau_{y1} - \tau_y$, for any $y \in \mathcal{N}$,
- (c) and $\sigma_{yi} := \tau_{yi} - \tau_{y(i-1)}$, for each $y \in \mathcal{N}$ and $i \geq 2$, $i \in \mathbb{I}$.

Let the function $\widehat{\varrho} : (0, \infty) \rightarrow (0, \infty)$ be defined as

$$\widehat{\varrho}(\lambda) := \mathbf{E} \sum_{j=1}^K e^{-\lambda \tau_j} = \sum_{j=1}^K \prod_{i=0}^{j-1} \frac{w(i)}{\lambda + w(i)} . \quad (2)$$

The function $\widehat{\varrho}$ plays a central role in the theory of the branching processes related to our model, as discussed in [21].¹ However, in the present work we use little of that relation – instead, we list here the known results that we will use.

1. The equation

$$\widehat{\varrho}(\lambda) = 1$$

has a unique root $\lambda^* > 0$. This λ^* is called the Malthusian parameter.

2. This λ^* gives the rate of exponential growth of the tree size almost surely. The normalized size of the tree converges almost surely to a random variable, which we denote by

$$\Theta := \lim_{t \rightarrow \infty} e^{-\lambda^* t} |\Upsilon(t)| .$$

3. Θ is almost surely positive, and

$$0 < \mathbf{E}\Theta < \infty, \quad (3)$$

which implies (also) that almost surely the process $\Upsilon(t)$ does not blow up in finite time.

¹The reason for the notation $\widehat{\varrho}$ is that this function is the Laplace transform of the density of the point process formed by birth times in the first generation of the tree.

4. Moreover,

$$\mathbf{E}\Theta^2 < \infty. \quad (4)$$

The first statement is in our setting obvious from the definition, since we have assumed $2 \leq K < \infty$. The second and third are shown in [21]. The last statement is also implicit from [21] – the variance is even calculated. Alternatively, the finiteness of the variance follows from Theorem 6.8.1 in [13], which states L^2 convergence of the normalized size under the condition $\mathbf{E}[(\sum_{i=1}^K e^{-\lambda\tau_i})^2] < \infty$, which is again obvious, since $K < \infty$.

Remark 2.1. *The process $\Upsilon(t)$ has an alternative construction, which we state here and refer to later. Define a countably infinite number of independent random variables $\tilde{\sigma}_x$, indexed with the elements of \mathcal{N} , as follows. Let $\tilde{\sigma}_\emptyset = 0$, and for $x = i_1 i_2 \dots i_n$, let $\tilde{\sigma}_x$ be exponentially distributed with parameter $w(i_n - 1)$. Denoting the parent of x by $p(x)$, we define $\tilde{\tau}_\emptyset = 0$ and*

$$\tilde{\tau}_x = \tilde{\tau}_{p(x)} + \tilde{\sigma}_{p(x)1} + \tilde{\sigma}_{p(x)2} + \dots + \tilde{\sigma}_{p(x)i_n}.$$

It is straightforward that with $\tilde{\Upsilon}(t) := \{x \in \mathcal{N} : \tilde{\tau}_x \leq t\}$, the process $\tilde{\Upsilon}$ has the same distribution as Υ .

2.3 Limiting Objects

Let $\Upsilon_{\downarrow x}(t) = (\Upsilon(t))_{\downarrow x}$ denote the subtree of $\Upsilon(t)$ rooted at x , which is the set of descendants of x (including x) that are born up to time t . (Note that t here is total time, and not the time since birth of x . In particular, $|\Upsilon_{\downarrow x}(0)| = 0$ if x is not the root.) For every $x \in \mathcal{N}$, we introduce the variables Θ_x , corresponding to the growth of the subtree under x , analogously to Θ ,

$$\Theta_x := \lim_{t \rightarrow \infty} e^{-\lambda^*(t-\tau_x)} |\Upsilon_{\downarrow x}(t)|.$$

The letter Θ refers to the variable corresponding to the root. Clearly, for every $x \in \mathcal{N}$, the random variables Θ_x are identically distributed. The basic relation between the different Θ_x variables in the tree is that for any $x \in \mathcal{N}$,

$$\Theta_x = \sum_{i=1}^K e^{-\lambda^*(\tau_{xi}-\tau_x)} \Theta_{xi}, \quad (5)$$

which is straightforward from $|\Upsilon_{\downarrow x}(t)| = 1 + \sum_{i=1}^K |\Upsilon_{\downarrow xi}(t)|$.

Now let us ask the following question. Fix a vertex $x \in \mathcal{N}$, and at time t , draw a vertex ζ_t uniformly randomly from $\Upsilon(t)$. What is the probability that ζ_t is descendant of x , so $x \prec \zeta_t$? As shown in (6) below, this probability tends to an almost sure limit Δ_x as $t \rightarrow \infty$, which can be expressed using the τ and Θ random variables,

$$\Delta_x := \lim_{t \rightarrow \infty} \frac{|\Upsilon_{\downarrow x}(t)|}{|\Upsilon(t)|} = e^{-\lambda^*\tau_x} \lim_{t \rightarrow \infty} \frac{e^{-\lambda^*(t-\tau_x)} |\Upsilon_{\downarrow x}(t)|}{e^{-\lambda^*t} |\Upsilon(t)|} = \frac{e^{-\lambda^*\tau_x} \Theta_x}{\Theta_\emptyset}. \quad (6)$$

We can now, for any $n \in \mathbb{N}$, define a random measure μ_n on the finite set $\{x : |x| = n\}$ (on the n^{th} generation of the tree), by

$$\mu_n(\{x\}) := \Delta_x.$$

This is a probability measure almost surely, which follows from the facts $\Delta_\emptyset = 1$ and $\Delta_y = \sum_{i=1}^K \Delta_{yi}$.

Let H_n denote the entropy of μ_n , that is

$$H_n = - \sum_{|x|=n} \Delta_x \log \Delta_x.$$

2.3.1 A Measure as the Limiting Object for the Tree

Let $\partial\mathcal{N}$ denote the set of leaves of the complete tree: $\partial\mathcal{N} = \{1, 2, \dots, K\}^\infty$. The concatenation xy makes sense for $x \in \mathcal{N}$ and $y \in \partial\mathcal{N}$, and then $xy \in \partial\mathcal{N}$. Also, for $x \in \mathcal{N}$ and $z \in \partial\mathcal{N}$, we write $x \prec z$ if $\exists y \in \partial\mathcal{N}$ such that $z = xy$. For $x \in \mathcal{N}$ we denote the set of leaves under x by $\partial\mathcal{N}(x) = \{z \in \partial\mathcal{N} : x \prec z\}$.

Let $\partial\mathcal{N}$ be equipped with the usual metric

$$d(x, y) = \Lambda^{\max\{n \in \mathbb{N} : x_{|n} = y_{|n}\}}, \quad (7)$$

where $0 < \Lambda < 1$ is an arbitrary constant. This constant is often chosen to be $1/e$, which makes certain formulae appear simpler. Yet we will not fix the value, so that our formulae express the dependence of the studied quantities on this arbitrary choice.

With the help of the μ_n random limiting measures, we define μ on the cylinder sets $\partial\mathcal{N}(x)$ of $\partial\mathcal{N}$ by

$$\mu(\partial\mathcal{N}(x)) := \mu_n(\{x\}) = \Delta_x, \text{ if } |x| = n,$$

and then we extend μ from $\{\partial\mathcal{N}(x) : x \in \mathcal{N}\}$ to the sigma-algebra generated (on $\partial\mathcal{N}$). Our results concern the properties of this extended random measure μ .

Remark 2.2. *Now we can tell why we use the continuous and not the discrete-time model in our work. The limiting relative weights Δ_x defined in (6) also make sense and are interesting in the discrete-time setting, just like the measure μ and the entropy H_n . Our results are formulated in terms of these quantities. However, the limiting “absolute” weights Θ_x , which will play a central role in the proofs, don’t make sense in the discrete-time setting.*

2.3.2 Dimensions of Measures: Definitions

For the reader’s convenience, let us review the definitions of local dimension, Hausdorff dimension and packing dimension of measures. The lower and upper local dimensions of μ at x are defined in [9] (2.15) and (2.16) as

$$\underline{\dim}_{\text{loc}} \mu(x) = \liminf_{r \rightarrow 0} \frac{\log \mu(B(x, r))}{\log r}, \quad (8)$$

$$\overline{\dim}_{\text{loc}} \mu(x) = \limsup_{r \rightarrow 0} \frac{\log \mu(B(x, r))}{\log r}, \quad (9)$$

where $B(x, r)$ is the ball of radius r centred at x . If the lower and upper local dimensions coincide at some x , they are called the local dimension at x . The Hausdorff and packing dimensions of μ are defined in [9] (10.8) and (10.9) as

$$\dim_{\text{H}} \mu = \sup\{s : \underline{\dim}_{\text{loc}} \mu(x) \geq s \text{ for } \mu\text{-almost all } x\}, \quad (10)$$

$$\dim_{\text{P}} \mu = \sup\{s : \overline{\dim}_{\text{loc}} \mu(x) \geq s \text{ for } \mu\text{-almost all } x\}. \quad (11)$$

The name of these dimensions come from the fact ([9] (10.10) and (10.11)) that

$$\dim_{\text{H}} \mu = \inf\{\dim_{\text{H}} E : E \text{ is a Borel set with } \mu(E) > 0\},$$

$$\dim_{\text{P}} \mu = \inf\{\dim_{\text{P}} E : E \text{ is a Borel set with } \mu(E) > 0\}.$$

We are ready to state our results.

2.4 Results

Theorem 2.3. *The limiting entropy*

$$h := \lim_{n \rightarrow \infty} \frac{1}{n} H_n$$

exists and is constant with probability one.

Theorem 2.4. *The Hausdorff dimension $\dim_H \mu$ and the packing dimension $\dim_P \mu$ of the measure μ are constant and equal with probability one, and h and the dimensions satisfy the relation*

$$\dim_H \mu = \dim_P \mu = \frac{h}{-\log \Lambda}, \quad (12)$$

where Λ is from (7). Moreover, the local dimension of μ equals $\dim_H \mu = \dim_P \mu$ at μ -almost every point.

Theorem 2.5. *Furthermore, an explicit formula for h is given:*

$$h = \mathbf{E} \left(\sum_{i=1}^K \lambda^* \tau_i e^{-\lambda^* \tau_i} \right).$$

This can be computed given the weight function w .

2.5 Some Related Models and Results

In the last decades there has been much progress in describing the asymptotic structure of randomly evolving trees, especially tree growth processes based on fragmentation processes. These processes are closely related to our model, see Remark 3.6. Limiting objects called “random real trees” and “continuum random trees” were introduced, to which the evolving trees converge, after an appropriate rescaling of the *distances* on the tree. Much of the structure of these limiting objects is understood, see e.g. [10, 12, 11].

Our concept of the limiting measure μ is different from these. It is a measure on the set of leaves of the infinite complete tree (with each vertex having exactly K children), which is a metric space, but the metric structure is trivial: it is not a result of any spatial scaling, and it carries no information about the tree growth process. On the other hand, the weights given by μ are a result of an appropriate rescaling of the tree size, where size means *cardinality*. In short, we are really interested in the asymptotic weight distribution, and not the asymptotic metric structure. This asymptotic weight distribution is also studied in the Physics literature, see e.g. [2], where a quantity analogous to the local dimension is calculated for a continuous time fragmentation process.

Population growth models, studied excessively in the theory of branching processes (see e.g. [13]), are also intimately related to our model, as discussed in detail in [21]. Scientists discussed the Hausdorff dimension of the *set* of individuals that are actually (sooner or later) born. However, in our model this is uninteresting, because – almost surely – every vertex is eventually born. Indeed, it is not the set, but the measure which captures the long-term structure of the tree well, and of which the dimension is interesting.

Similarly, in the limiting continuous trees obtained in [10, 12, 11] by a spatial rescaling of the evolving tree, the metric structure is of main interest, and the Hausdorff dimension and Hausdorff measure of *sets* are the natural questions to ask [8, 7] – unlike in our setting.

The continuous time version of our tree growth process can also be translated into a branching random walk, with time turning into displacement. Then the asymptotic growth can be described analogously, see the Biggins theorem in [4] or [16]. However, with that point of view, the natural questions about the limiting structure are quite different.

3 Main Line of the Proof

3.1 Idea of the Proof

The random limiting measure μ depends on the random growth of the tree. The idea of the proof is the following: we define a random leaf in the limiting tree according to the measure μ . The way the random leaf is defined is based on a step-by-step construction of the subsequent generations of the limiting tree, together with a step-by-step construction of a path from the root to the random leaf. This is done in such a way that a Markov process appears naturally along this path, and the local dimension of the measure μ in this random point can be computed as an ergodic average. It follows

that this average is constant with probability one, unconditionally. Thus, although the measure depends on the random tree growth, this ergodic average is constant, and it is the local dimension of the measure in all the μ -typical leafs of the limiting tree. This implies that this constant is the Hausdorff (and also the packing) dimension of μ with probability one. Some technical difficulty comes from the fact that the state space of the key Markov process is continuous and non-compact, so to apply ergodic theorems, one has to work for the existence of the invariant measure (while uniqueness is easy).

3.2 Markov Structure of the Tree

The content of this short section is mainly repetition of material from [20]. These concepts and statements allow for a good understanding of the tree structure, on which our main construction (in Section 3.3) relies. Lemma 3.2 will also be used formally in Section 3.3 to get an easy proof of the fact that our step-by-step construction of the limiting tree is equivalent to the original model (Proposition 3.7).

Definition 3.1. *We say that a system of random variables $(Y_x)_{x \in \mathcal{N}}$ constitutes a tree-indexed Markov field if for any $x \in \mathcal{N}$, the distribution of the collection of variables $(Y_y : x \prec y)$, and that of $(Y_z : x \not\prec z)$, are conditionally independent, given Y_x .*

We state the following:

Lemma 3.2. *For each $x \in \mathcal{N}$ let V_x denote the vector $V_x := (\sigma_x, \Theta_x)$. Then the collections of variables $\mathcal{A}_x := (V_y : x \prec y)$ and $\mathcal{B}_x := (V_z : x \not\prec z; \sigma_x)$ are conditionally independent, given Θ_x .*

Proof. Recall Remark 2.1, the alternative construction of $\Upsilon(t)$. From that, it is straightforward that the collection \mathcal{A}_x is in fact constructed by the set of independent variables $A_x := (\sigma_y : x \prec y)$.

Similarly, recall (5), and decompose $\Theta_{p(x)}$, where $p(x)$ is the parent of vertex x ,

$$\Theta_{p(x)} = \sum_{j=1}^K e^{-\lambda^*(\tau_{p(x)j} - \tau_{p(x)})} \Theta_{p(x)j} = \sum_{j=1}^K e^{-\lambda^*(\sigma_{p(x)1} + \sigma_{p(x)2} + \dots + \sigma_{p(x)j})} \Theta_{p(x)j}.$$

This means that if we take the set of variables $B_x := (\sigma_y : x \not\prec y)$, then \mathcal{B}_x is constructed by $B_x \cup \{\Theta_x\}$.

Given Θ_x , the two collections $\mathcal{A}_x \cup \{\Theta_x\}$ and $\mathcal{B}_x \cup \{\Theta_x\}$ are conditionally independent, this way the same is true for \mathcal{A}_x and \mathcal{B}_x , so the statement of the lemma follows. \square

Corollary 3.3. *The variables $(\Theta_x)_{x \in \mathcal{N}}$ constitute a tree-indexed Markov field.*

Proof. Direct consequence of Lemma 3.2, since $V_x = (\sigma_x, \Theta_x)$. \square

Definition 3.4. *We introduce the variables R_x , indexed by \mathcal{N} . For the root we leave R_\emptyset undefined. For any other vertex y' which has a parent y , so for any $y' = yi$ with $i \in \mathbb{I}$, let*

$$R_{yi} := \lim_{t \rightarrow \infty} \frac{|\Upsilon_{\downarrow yi}(t)|}{|\Upsilon_{\downarrow y}(t)|} = \frac{e^{-\lambda^*(\tau_{yi} - \tau_y)} \Theta_{yi}}{\Theta_y} = \frac{\Delta_{yi}}{\Delta_y}.$$

Notice that for $x = (i_1 i_2 \dots i_n)$, Δ_x is a telescopic product,

$$\Delta_x = \Delta_{i_1} \frac{\Delta_{i_1 i_2}}{\Delta_{i_1}} \frac{\Delta_{i_1 i_2 i_3}}{\Delta_{i_1 i_2}} \dots \frac{\Delta_{i_1 \dots i_n}}{\Delta_{i_1 \dots i_{n-1}}} = R_{i_1} R_{i_1 i_2} R_{i_1 i_2 i_3} \dots R_{i_1 \dots i_n}.$$

Equivalently, for $|x| = n$,

$$\log \Delta_x = \sum_{l=1}^n \log R_{x|_l}, \quad (13)$$

where $x|_l$ denotes the first l letters of the string x (which denotes the ancestor of x on the l -th level of the tree).

3.3 Construction of the Random Leaf

We will now give a different construction of the tree from the ones seen before. Namely, we construct the system of $V_x = (\sigma_x, \Theta_x)$ variables starting from the root, and going step-by-step, from generation to generation. Together with these, we compute the R_x and Δ_x variables, and use them to construct a random path $\{y_n\}$ from the root to the edge of the infinite tree. The y_n will be chosen from the children of y_{n-1} in a “size-biased” way. We will use this path in the proofs of our results. For the sake of simple notation, we suppose for a moment that the maximum number of children of any vertex is two, that is, $K = 2$. It is straightforward to construct the corresponding generations and the random path for any $K < \infty$. For the rest of this section we treat the distribution of Θ as known.

Recall that $\sigma_1, \sigma_2, \Theta_1$ and Θ_2 are independent. Keeping that in mind, using

$$\Theta = e^{-\lambda^* \sigma_1} (\Theta_1 + e^{-\lambda^* \sigma_2} \Theta_2), \quad (14)$$

we will consider the conditional joint distribution of $(\sigma_2, \Theta_1, \Theta_2)$, given Θ . (Of course, σ_1 is – conditionally – a deterministic function of these, but we will not use the value.) Now we can construct the generations, together with the random path y_n , in the following steps.

1. Pick Θ_\emptyset at random, according to its distribution, and fix $\sigma_\emptyset = 0$. Also, fix $y_0 = \emptyset$.
2. *First generation*
 - (a) Pick $(\sigma_2, \Theta_1, \Theta_2)$ according to their conditional distribution, given Θ_\emptyset
 - (b) Define $\Delta_1 = R_1 = \frac{\Theta_1}{\Theta_1 + e^{-\lambda^* \sigma_2} \Theta_2}$ (which is equal to $\frac{e^{-\lambda^* \sigma_1} \Theta_1}{\Theta}$, and happens not to depend on σ_1). Also define $\Delta_2 = R_2 = \frac{e^{-\lambda^* \sigma_2} \Theta_2}{\Theta_1 + e^{-\lambda^* \sigma_2} \Theta_2}$.
 - (c) Choose y_1 according to the conditional probabilities $\mathbf{P}(y_1 = 1 | \Theta, \sigma_2, \Theta_1, \Theta_2) = R_1$ and $\mathbf{P}(y_1 = 2 | \Theta, \sigma_2, \Theta_1, \Theta_2) = R_2$.
3. *Second generation*
 - (a) Repeat the steps seen before for the progeny of vertex 1, to get $(\sigma_{12}, \Theta_{11}, \Theta_{12})$ and also R_{11} and R_{12} . This is done only using the information carried by Θ_1 , conditionally independently of (Θ, Θ_2) . This conditional independence is the consequence of Corollary 3.3. Since we already know R_1 , we can now compute the values $\Delta_{11} = R_1 R_{11}$ and $\Delta_{12} = R_1 R_{12}$.
 - (b) Independently of the previous steps, use Θ_2 to get $(\sigma_{22}, \Theta_{21}, \Theta_{22})$, R_{21} and R_{22} . We then also have Δ_{21} and Δ_{22} .
 - (c) Choose y_2 from the children of y_1 , according to the conditional distribution given by the R_x variables in the second generation. Namely, if $y_1 = 1$,

$$\mathbf{P}(y_2 = 11 | y_1 = 1, \sigma_{12}, \Theta_{11}, \Theta_{12}) = R_{11}$$

$$\mathbf{P}(y_2 = 12 | y_1 = 1, \sigma_{12}, \Theta_{11}, \Theta_{12}) = R_{12},$$

and if $y_1 = 2$,

$$\mathbf{P}(y_2 = 21 | y_1 = 2, \sigma_{22}, \Theta_{21}, \Theta_{22}) = R_{21}$$

$$\mathbf{P}(y_2 = 22 | y_1 = 2, \sigma_{22}, \Theta_{21}, \Theta_{22}) = R_{22},$$

conditionally independently of the entire past of the construction.

4. n -th generation

- (a) Having constructed all the Θ_x with $|x| = n-1$, split these all in the way above, conditionally independently of each other (and the entire past of the construction), to get the R_z and Δ_z variables in the n -th generation. In particular,

$$R_{xi} = \frac{e^{-\lambda^* (\sigma_{x1} + \dots + \sigma_{xi})} \Theta_{xi}}{\Theta_x}.$$

- (b) According to the value of y_{n-1} , choose y_n from its children, according to the corresponding R_z distribution (conditionally independently of the entire past).

Remark 3.5. As mentioned before, our model is intimately related to a branching process, as discussed in [21]. In branching processes, the idea of size biasing is not at all new, as its importance is emphasized e.g. in [17].

Remark 3.6. This step-by-step construction of the tree is similar to the fragmentation processes discussed e.g. in [3]. There the usage of “randomly tagged branches” based on size-biased choices is a standard technique, see [3], Section 1.2.3. Note however, that our step-by-step construction is not a fragmentation process in the classical sense. In particular, the sequence of measures μ_n is not Markov: the process also “remembers” the values Θ_x which influence how the weight $\mu_n(\{x\})$ at x is further “fragmented”.

Proposition 3.7. With $V_x = (\sigma_x, \Theta_x)$ as before, the distribution of $\{V_x\}_{x \in \mathcal{N}}$ in the above construction is identical to the distribution in the randomly growing tree model.

Proof. The statement we are proving is about the joint distribution of countably infinitely many (real-valued) random variables, so this joint distribution can be viewed as a measure on $\mathbb{R}^{\mathbb{N}}$,² with the σ -algebra of measurable sets being the σ -algebra generated by cylinder sets – defined in terms of finitely many of the σ_x and Θ_x . So to prove that the two measures on $\mathbb{R}^{\mathbb{N}}$ – given by the two constructions – coincide, it is enough to see that they coincide on such cylinder sets.

In terms of joint distributions: It is enough to see that the distributions of $\{V_x\}_{x \in \mathcal{N}}$ coming from the two constructions have identical finite-dimensional marginals. In particular, it is enough to show that for every n , the distribution of $\{V_x\}_{x \in \mathcal{N}, |x| \leq n}$ in the above construction is identical to the distribution in the randomly growing tree model.

This is easy to see by induction:

- For $n = 0$ we have chosen the law of Θ_\emptyset properly by construction, also $\sigma_\emptyset = 0$ as it should be.
- For $n = 1$, the $\{V_x\}_{x \in \mathcal{N}, |x|=1}$ are constructed to have the right conditional joint distribution, given Θ_\emptyset , so the $n = 0$ statement implies the $n = 1$ statement. In particular, the Θ_x for $|x| = 1$ are distributed as they should be.
- For $n \geq 2$, the same argument (the construction) gives inductively that the joint distribution of the $\{V_x\}_{x \in W}$ is what it should be, for any family W of x -es which consists of a vertex and its children. However, the construction also ensures the conditional independence of $\{V_y\}_{x \prec y}$ and $\{V_z\}_{x \not\prec z}$ given Θ_x , as in Lemma 3.2. This, together with the joint distributions of the $\{V_x\}_{x \in W}$ (with W as above) already characterizes the joint distribution of $\{V_x\}_{x \in \mathcal{N}, |x| \leq n}$.

□

From now on, we will use the alternative construction of the tree in our discussion, so Proposition 3.7 is used all the time in the proof, but this will not be formally mentioned.

Definition 3.8. Denote by Υ the σ -algebra generated by $\{\sigma_x \mid x \in \mathcal{N}\}$, which contains the full tree evolution.

Note that for any $x \in \mathcal{N}$, Θ_x is measurable with respect to Υ , so Υ is also the σ -algebra generated by $\{\sigma_x, \Theta_x \mid x \in \mathcal{N}\}$, namely all the data about the tree – but not about the random leaf – during the parallel construction of the tree and the random leaf just presented.

The usefulness of the random leaf we constructed is shown by the following:

Lemma 3.9. Conditioned on Υ , the conditional distribution of the leaf $\lim_n y_n$ is exactly the measure μ . Similarly, the conditional distribution of y_n is exactly μ_n .

Proof. The second statement can be seen by induction: μ_0 obviously gives weight 1 to the single point $\emptyset = y_0$. Later, by construction of y_{n+1} , for any $x \in \mathcal{N}$ with $|x| = n$ and any $i \in \mathbb{I}$ we have $\mathbf{P}(y_{n+1} = xi \mid y_n = x, \Upsilon) = R_{xi}$, so if we assume inductively that $\mathbf{P}(y_n = x \mid \Upsilon) = \mu_n(\{x\}) = \Delta_x$, then $\mathbf{P}(y_{n+1} = xi \mid \Upsilon) = \Delta_x R_{xi} = \Delta_{xi} = \mu_{n+1}(\{xi\})$ for any $|xi| = n + 1$, so y_{n+1} is indeed distributed according to μ_{n+1} .

The first statement is an immediate consequence of the second, since for any cylinder set $\partial\mathcal{N}(x)$, if $|x| = n$, we have $\mathbf{P}(y_\infty \in \partial\mathcal{N}(x) \mid \Upsilon) = \mathbf{P}(y_n = x \mid \Upsilon) = \mu_n(\{x\}) = \mu(\partial\mathcal{N}(x))$. □

²we could write $([0, \infty) \times [0, \infty))^{\mathbb{N}}$, but a measure on this can be viewed as a special case of a measure on $\mathbb{R}^{\mathbb{N}}$.

Corollary 3.10. *Conditioned on the tree, the conditional expectation of $-\log \Delta_{y_n}$ is exactly H_n .*

Proof. Indeed, by the above lemma,

$$\mathbf{E}(-\log \Delta_{y_n} | \Upsilon) = - \sum_{|x|=n} \mathbf{P}(y_n = x | \Upsilon) \log \Delta_x = - \sum_{|x|=n} \mu_n(\{x\}) \log \Delta_x = - \sum_{|x|=n} \Delta_x \log \Delta_x = H_n.$$

□

3.4 Markov Processes Along the Random Path

The key to the proof is the following easy observation:

Proposition 3.11. *The stochastic process $X_n = \Theta_{y_n}$ ($n = 0, 1, 2, \dots$) is a homogeneous Markov process. By “homogeneous” we mean that the transition kernel does not depend on n .*

Proof. This is clear from the construction in Section 3.3. Indeed, when constructing Θ_{y_n} , only the value of $\Theta_{y_{n-1}}$ is used, and the construction is the same on every level. □

The reason to construct in Section 3.3 the entire tree of pairs (Θ_x, Δ_x) step by step – and not just the random path $\{y_n\}$ on an already existing tree – was exactly to make the Markov property of Θ_{y_n} obvious. A direct proof without the step-by-step construction would also not be hard, but according to our taste, the underlying phenomena are more transparent this way.

Based on this proposition and equation (13), the proof of our main results will be a reference to an appropriate ergodic theorem. However, there are two issues to deal with before. First, the state space of our Markov processes is continuous and even non-compact, so the unique existence of the invariant measure needs to be discussed. This is done in the next proposition. Second, the quantity $-\log R_{y_n}$, of which we want to calculate the ergodic average, is not an observable on the state space of X_n , so this state space needs to be extended. This obvious extension will be done in Corollary 3.16.

Before starting the main arguments, let us formulate, as a lemma, an easy observation about the distribution of Θ . We will use this in the arguments both for the uniqueness and the existence of the invariant measure of X_n . From now on, we will use the notation \mathbb{R}^+ for the set of *positive* real numbers:

$$\mathbb{R}^+ = (0, \infty).$$

It is important that 0 is not included, e.g. when we speak of functions being continuous or nonzero on \mathbb{R}^+ .

Lemma 3.12. *Θ is absolutely continuous w.r.t. Lebesgue measure on \mathbb{R}^+ , with a density function π which is continuous and strictly positive on \mathbb{R}^+ .*

Proof. Start from the decomposition (5). It shows that Θ is of the form $\Theta = e^{-\lambda^* \sigma_1} \widehat{\Theta}$ where σ_1 is independent of $\widehat{\Theta}$, which immediately implies that Θ must be equivalent to Lebesgue measure on the interval from zero to its maximal value. On the other hand, $\Theta \geq e^{-\lambda^* \sigma_1} \Theta_1 + e^{-\lambda^* (\sigma_1 + \sigma_2)} \Theta_2$ implies that Θ is not bounded, since Θ_1 and Θ_2 are independent and distributed as Θ , and their prefactors can be arbitrarily close to 1. The same decomposition, applied once again, also implies that the density π is even a continuous function (more precisely, can be chosen to be continuous), since Θ being absolutely continuous w.r.t. Lebesgue measure implies that so is $\widehat{\Theta}$ (since $K < \infty$), the density of which is once again smoothened by $\Theta = e^{-\lambda^* \sigma_1} \widehat{\Theta}$. □

For the discussion of the invariant measures, let P denote the transition kernel of X_n – that is, $P(t)$ is the conditional distribution of X_{n+1} under the condition $X_n = t$ (for every $t \in \mathbb{R}^+$). We also use it as the operator acting on measures by $\eta P = \int_{\mathbb{R}^+} P(t) d\eta(t)$.

Proposition 3.13. *The transition kernel P of the Markov process $X_n = \Theta_{y_n}$ has exactly one invariant measure.*

Proof. Recall that the decomposition (5) is the key relation between the Θ_x -es of the different generations, on which the construction of X_n – and thus every property of the transition kernel – is based.

The key observation is that $P(t)$ is equivalent to Lebesgue measure (on \mathbb{R}^+ , of course) for every $t \in \mathbb{R}^+$. This (and more) is explicitly stated and proven in Lemma 4.5. However, since we feel that this statement is really intuitive, let us give a rough reasoning here as well.

First, Lemma 3.12 implies that the distribution of Θ is equivalent to Lebesgue measure on \mathbb{R}^+ . Recall now the construction in Section 3.3, the essence of which is that $P(t)$ is the conditional distribution of Θ' under the condition $\Theta = t$, where Θ' is a random choice from the set $\{\Theta_1, \dots, \Theta_K\}$. Look again at the relation between Θ and $\{\Theta_1, \dots, \Theta_K\}$, which is the decomposition (5), or the simplified form for $K = 2$, which is (14). It shows that given any value of t , the condition $\Theta = t$ doesn't rule out any of the possible values of a Θ_i with $1 \leq i \leq K$. Also, the conditioning on $\Theta = t$ doesn't spoil the absolute continuity of Θ_i , and the method of randomly choosing Θ' from $\{\Theta_1, \dots, \Theta_K\}$ also preserves absolute continuity. With this, the key observation is shown. Again, see Lemma 4.5 for a detailed proof.

This observation about $P(t)$ implies that for any measure η on \mathbb{R}^+ , the first iterate ηP is already equivalent to Lebesgue measure. This in turn implies that any invariant measure $\eta = \eta P$ is equivalent to Lebesgue measure, so any two invariant measures are equivalent.

Suppose now indirectly that there exist two different invariant probability measures. Then two different *extremal* invariant probability measures also have to exist. But two different extremal invariant probability measures must be mutually singular, which contradicts the previous argument. Thus there is at most one invariant probability measure.

The existence follows from Lemma 3.15 and Lemma 3.14. Indeed, the limiting measure ν of Lemma 3.14 has to be invariant by Lemma 3.15. \square

Lemma 3.14. *The sequence of random variables $X_n = \Theta_{y_n}$ is weakly convergent to some measure ν on \mathbb{R}^+ .*

To keep our arguments easy to follow, we delay the proof to Section 4.2.

Lemma 3.15. *P is continuous with respect to weak convergence of measures.*

The proof is delayed to Section 4.3.

Corollary 3.16. *The stochastic process $Y_n = (\Theta_{y_n}, R_{y_n})$ ($n = 1, 2, \dots$) is a homogeneous Markov process, for which the transition kernel has exactly one invariant measure.*

Proof. Notice that during the construction of the tree in Section 3.3, R_{y_n} is constructed by using only the value of $\Theta_{y_{n-1}}$ (not even $R_{y_{n-1}}$), in a time-homogeneous way. Thus Y_n is really homogeneous Markov. Let \tilde{P} denote the transition kernel. From the construction, $\tilde{\eta}\tilde{P}$ depends only on the first marginal of $\tilde{\eta}$, and on this marginal it acts exactly like P . So for any measure $\tilde{\nu}$ with first marginal ν , $\tilde{\nu} := \tilde{\nu}\tilde{P}$ is invariant by the invariance of ν under P . The uniqueness is obvious from the uniqueness of ν . \square

Now we are ready to apply an ergodic theorem on the sequence $-\log R_{y_n}$ to get the central technical result, from which our first two theorems easily follow.

Corollary 3.17. *The limit $h := -\lim_{n \rightarrow \infty} \frac{1}{n} \log \Delta_{y_n}$ exists and is constant with probability one.*

Proof. $-\log R_{y_n}$ is an observable on the state space of Y_n , and h is exactly the ergodic average of this observable by (13). So it is guaranteed to be constant by the unique existence of the invariant measure and Theorem 1.1 in Chapter X of [6]. We give the details of the (standard) argument now.

Theorem 1.1 in Chapter X of [6] states that “If $\{x_n, n \geq 0\}$ is a stationary Markov process, and if z is an invariant random variable, then z is measurable on the sample space of x_0 ”. To formally apply this theorem to our process, we first need to construct a stationary version of Y_n . Namely, let \tilde{Y}_n be the Markov process with generator \tilde{P} started from \tilde{Y}_0 which is distributed according to the unique invariant measure $\tilde{\nu}$. For this process, the ergodic average of an observable, being an invariant random variable (see [6], Chapter X for the definition), is by the above theorem measurable

on the state space – that is, constant with probability one, conditioned on the initial value (more precisely, for $\tilde{\nu}$ -a.e. initial value). But in our case, this constant is indeed independent of the initial value – actually, it is constant for *every* initial value, since \tilde{P} brings any measure (e.g. a point measure concentrated on any point) into a measure equivalent with $\tilde{\nu}$. Now notice that the property that the ergodic average is the same constant with probability one, independently of the initial state, is a property of the transition kernel \tilde{P} only (and not of \tilde{Y}_n as a stochastic process), so it also holds for the process Y_n . \square

Remember that $\frac{1}{n}H_n$ is a conditional expectation of $-\frac{1}{n}\log \Delta_{y_n}$ by Corollary 3.10. So since we have just shown the almost sure convergence of $-\frac{1}{n}\log \Delta_{y_n}$, the almost sure convergence of $\frac{1}{n}H_n$ follows, if we have e.g. dominated convergence. This will be guaranteed by the following lemma.

Lemma 3.18. *Let $\tilde{\mu}$ be any Borel probability measure on $\partial\mathcal{N}$, with $K < \infty$. Using the notation in Section 2.3.1, for every $x \in \partial\mathcal{N}$ let*

$$f_n(x) := -\frac{1}{n}\log \tilde{\mu}(\partial\mathcal{N}(x|_n)).$$

Then $\bar{f} := \sup_n f_n$ is integrable with respect to the measure $\tilde{\mu}$.

The proof is delayed to Section 4.1. Now we are ready to prove the main results of the paper.

Proof of Theorem 2.3. For every $x \in \partial\mathcal{N}$ let $f_n(x) = -\frac{1}{n}\log \mu_n(\{x|_n\}) = -\frac{1}{n}\log \mu(\partial\mathcal{N}(x|_n))$. By Lemma 3.9, Corollary 3.17 states exactly that for almost every realization of the tree, $f_n(x)$ converges μ -almost surely to h .

Now divide the statement of Corollary 3.10 by n to get

$$\frac{1}{n}H_n = \mathbf{E}\left(-\frac{1}{n}\log \Delta_{y_n} | \Upsilon\right) = \int_{\{\bar{x} \in \mathcal{N} : |\bar{x}|=n\}} -\frac{1}{n}\log(\mu_n(\{\bar{x}\})) d\mu_n(\bar{x}) = \int_{\partial\mathcal{N}} f_n(x) d\mu(x).$$

We can now apply the dominated convergence theorem to finish the proof, since we can use the supremum as an integrable dominating function, see Lemma 3.18. \square

Proof of Theorem 2.4. We first show the second statement of the theorem by showing that the local dimension of μ at the leaf $\lim_n y_n$ is exactly $\frac{h}{-\log \Lambda}$ where h is from Corollary 3.17. Let $B(x, r)$ denote the r -neighbourhood of the point $x \in \partial\mathcal{N}$ w.r.t. the metric (7). For $r = \Lambda^n$, this neighbourhood is formed exactly by the descendants of $x|_n$, so $B(x, \Lambda^n) = \partial\mathcal{N}(x|_n)$. The μ -measure of this set is

$$\mu(B(x, \Lambda^n)) = \mu(\partial\mathcal{N}(x|_n)) = \mu_n(\{x|_n\}) = \log \Delta_{x|_n},$$

while the logarithm of the diameter of this set is $n \log \Lambda$. Thus the local dimension of μ at the leaf x is

$$\dim_{\text{loc}} \mu(x) = \lim_{n \rightarrow \infty} \frac{\mu(B(x, \Lambda^n))}{n \log \Lambda} = \lim_{n \rightarrow \infty} \frac{-\frac{1}{n} \log \Delta_{x|_n}}{-\log \Lambda}$$

(if this limit exists), by the definition in (8) and (9).

Applying that to $x = \lim_n y_n$, Lemma 3.9 and Corollary 3.17 say that this limit indeed exists and is equal to $\frac{h}{-\log \Lambda}$ for μ -almost every x , which is what we wanted to show.

The first statement of the theorem is now an immediate consequence of the definitions of the Hausdorff and packing dimension of a measure in (10) and (11). \square

4 Proofs of Auxiliary Lemmas

4.1 The Lemma for Dominated Convergence of the Entropies

In this section we prove Lemma 3.18.

Proof of Lemma 3.18. For arbitrary $M < \infty$, let us define the set

$$F_M^{(n)} := \{x : f_n(x) \geq M\} = \{x : -\frac{1}{n} \log \tilde{\mu}(\partial\mathcal{N}(x|_n)) \geq M\} = \{x : \tilde{\mu}(\partial\mathcal{N}(x|_n)) \leq e^{-nM}\}.$$

Since f_n takes constant values on the K^n cylinder sets, we have

$$\tilde{\mu}(F_M^{(n)}) \leq K^n e^{-nM} = (Ke^{-M})^n. \quad (15)$$

Now we define

$$F_M := \{x : \bar{f}(x) > M\} = \bigcup_n \{x : f_n(x) > M\} \subseteq \bigcup_n F_M^{(n)}.$$

By (15), for $M > \log(2K)$,

$$\tilde{\mu}(F_M) \leq \sum_{n=1}^{\infty} (Ke^{-M})^n < 2Ke^{-M}.$$

Thus, since $\bar{f} \geq 0$,

$$\int \bar{f}(x) d\tilde{\mu}(x) < \sum_{M=1}^{\infty} M \tilde{\mu}(\{x : M-1 \leq \bar{f}(x) < M\}) < \infty.$$

□

4.2 Limiting Distribution of Θ_{y_n} Along the Random Path

In this section we prove Lemma 3.14. We begin with three lemmas of elementary probability whose statements do not rely on the setting of the paper.

The first one is a trivial generalization of the ordinary weak law of large numbers. We could call it “Weak law of large numbers with arbitrary weights”. For this purpose, we will consider a sequence of probability vectors $\{\underline{p}^n\}_{n=1}^{\infty}$, where, again, each \underline{p}^n is a probability vector $\underline{p}^n = (p_1^n, p_2^n, \dots, p_{N_n}^n)$. We plan to calculate weighted averages of independent random variables with weight vectors \underline{p}^n . We expect such an average to be close to the expectation, if every term has a sufficiently small weight. So we will say that the sequence $\{\underline{p}^n\}_{n=1}^{\infty}$ is proper if

$$\lim_{n \rightarrow \infty} \max\{p_j^n : 1 \leq j \leq N_n\} = 0$$

.

Lemma 4.1. *Let ν_0 be a probability distribution on \mathbb{R} with finite expectation m . Let $\{\underline{p}^n\}_{n=1}^{\infty}$ be a proper sequence of weight vectors, and let ν_n be the distribution of*

$$\sum_{j=1}^{N_n} p_j^n Z_j$$

where Z_1, Z_2, \dots, Z_{N_n} are independent random variables with distribution ν_0 . Then

$$\nu_n \Rightarrow m.$$

Note that this is the usual weak law if $p_j^n = \frac{1}{n}$ ($j = 1, \dots, n$).

Proof. The proof is trivial following the standard proof of the weak law with characteristic functions. □

Now we turn to a lemma which could be called “size-biased sampling with arbitrary extra weights”. For this purpose, let $\underline{p} = (p_1, p_2, \dots, p_N)$ be a probability vector, and let Z_1, Z_2, \dots, Z_N be random variables on \mathbb{R}^+ (meaning $\mathbf{P}(Z_j > 0) = 1$). We will say that the random variable V is the size-biased random choice from Z_1, Z_2, \dots, Z_N with extra weights p_1, p_2, \dots, p_N , if it is constructed the following way:

1. Generate a realization of (Z_1, Z_2, \dots, Z_N) , and call it (z_1, z_2, \dots, z_N) .
2. Having that, choose a random integer J from the index set $\{1, 2, \dots, N\}$ with the weight

$$\frac{p_j z_j}{\sum_{j=1}^N p_j z_j}$$

given to each j .

3. Set $V = z_J$.

Note that this is the usual size-biased random choice if all the p_j are equal. Our lemma states that this size-biased random choice with extra weights behaves just like the ordinary one, provided that every weight is small.

To state the lemma, let ν_0 be a probability distribution on \mathbb{R}^+ with finite expectation m . We will say that the measure ν is the size-biased version of ν_0 , if it is absolutely continuous with respect to ν_0 , and the density is $\rho(t) = \frac{1}{m}t$. In other words, $\nu(A) = \frac{1}{m} \int_A t d\nu_0(t)$.

Lemma 4.2. *Let ν_0 be a probability distribution on \mathbb{R}^+ with finite expectation m . Let $\{\underline{p}^n\}_{n=1}^\infty$ be a proper sequence of weight vectors, and (for each n) let $Z_1^n, Z_2^n, \dots, Z_{N_n}^n$ be independent random variables with distribution ν_0 . Let V_n be the random choice from $Z_1^n, Z_2^n, \dots, Z_{N_n}^n$ with extra weights $p_1^n, p_2^n, \dots, p_{N_n}^n$. Let ν be the size-biased version of ν_0 . Then*

$$V_n \Rightarrow \nu.$$

Proof. Let F denote the cumulative distribution function of ν , that is, $F(t) = \nu([0, t])$. Let F_n denote the cumulative distribution function of V_n . For some fixed t , we write it in the form

$$F_n(t) = \mathbf{E}(\mathbf{P}(V_n \leq t \mid \{Z_j^n\}_{j=1}^{N_n})). \quad (16)$$

The conditional probability inside is just the weight of j -s with $Z_j \leq t$, so

$$\mathbf{P}(V_n \leq t \mid \{Z_j^n\}_{j=1}^{N_n}) = \frac{\sum_{j=1}^{N_n} p_j^n Z_j^n \mathbf{1}(Z_j^n \leq t)}{\sum_{j=1}^{N_n} p_j^n Z_j^n}.$$

According to Lemma 4.1 the denominator converges weakly (and thus, in probability) to $\mathbf{E}(Z_1^n) = m > 0$ as $n \rightarrow \infty$. Similarly, the numerator converges in probability to

$$\mathbf{E}(Z_1^n \mathbf{1}(Z_1^n \leq t)) = \int_{\mathbb{R}^+} \tilde{t} \mathbf{1}(\tilde{t} \leq t) d\nu_0(\tilde{t}) = m\nu([0, t]).$$

This implies that the quotient converges weakly to $\nu([0, t]) = F(t)$. Since this quotient is a conditional probability, it is obviously bounded by 1, so (16) implies that $F_n(t) \rightarrow F(t)$. \square

The following lemma is just a re-statement of the previous one. This is the form that we will use.

Lemma 4.3. *Let ν_0 be a probability distribution on \mathbb{R}^+ with finite expectation, and let ν be its size-biased version. Let ϕ be a bounded continuous function on \mathbb{R}^+ . Then for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for any probability vector (p_1, p_2, \dots, p_N) which satisfies that*

$$\max\{p_j : 1 \leq j \leq N\} \leq \delta,$$

if Z_1, Z_2, \dots, Z_N are independent with distribution ν_0 , then the size-biased random choice (called V) from Z_1, Z_2, \dots, Z_N with extra weights p_1, p_2, \dots, p_N satisfies

$$|\mathbf{E}(\phi(V)) - \int \phi(t) d\nu(t)| < \varepsilon.$$

Before proving Lemma 3.14, we need one more tiny statement about the structure of the growing tree.

Lemma 4.4. *For any vertex $x \in \mathcal{N}$, let*

$$T_x = e^{-\lambda^* \tau_x}, \quad (17)$$

and for every x with $|x| = n$ let

$$p_x = \frac{T_x}{\sum_{|y|=n} T_y}.$$

Then the sequence $p^{n,\max} := \max\{p_x : |x| = n\}$ converges to zero in probability.

Proof. We prove the stronger statement that $p^{n,\max}$ converges to zero with probability one. We use the form

$$p^{n,\max} = \frac{\max\{T_x : |x| = n\}}{\sum_{|y|=n} T_y}. \quad (18)$$

We show that the numerator converges to zero with probability one, while the denominator converges to a positive limit with probability one.

1. If the numerator does not converge to zero, then there is some $\varepsilon > 0$ and there are infinitely many vertices $x \in \mathcal{N}$ with $T_x > \varepsilon$. Then, for all these x we have $\tau_x < \tau^* := \frac{-\log \varepsilon}{\lambda^*}$, so infinitely many vertices are born within the finite time τ^* . This is known to have probability zero – see comment at (3).
2. Iterating the decomposition of Θ , we get

$$\Theta = \sum_{|x|=n} T_x \Theta_x. \quad (19)$$

Let Σ_n denote the σ -algebra generated by $\{\sigma_x : x \in \mathcal{N}, |x| \leq n\}$ – that is, the complete history of the tree growth up to the n -th level. Similarly, let Σ denote the σ -algebra generated by $\{\sigma_x : x \in \mathcal{N}\}$. Clearly $\Sigma_n \subset \Sigma_{n+1}$, Σ is generated by $\cup_n \Sigma_n$, and Θ is Σ -measurable. So Lévy's ‘upward’ theorem ensures that $\mathbf{E}(\Theta \mid \Sigma_n) \rightarrow \Theta$ with probability one. However, if $|x| = n$, then Θ_x is independent of Σ_n , while T_x is Σ_n -measurable, so (19) implies that

$$\mathbf{E}(\Theta \mid \Sigma_n) = \sum_{|x|=n} T_x \mathbf{E}\Theta_x = \mathbf{E}\Theta \sum_{|x|=n} T_x,$$

so with probability one the denominator of (18) converges to $\frac{\Theta}{\mathbf{E}\Theta} \neq 0$.

□

Now we can complete the goal of this subsection:

Proof of Lemma 3.14. Actually we give the limit explicitly. Let ν be the measure on \mathbb{R}^+ with density function $c x \pi(x)$, where $\pi(x)$ is the density of Θ , and $c = \frac{1}{\mathbf{E}\Theta}$ is a normalizing constant. We will show that

$$X_n \Rightarrow \nu. \quad (20)$$

Let us look directly at $X_n = \Theta_{y_n}$ for some fixed n . This can also be constructed in the following way:

1. Generate the birth times τ_x for all vertices x with $|x| = n$ (that is, on the n -th level of the tree). This defines the values $T_x = e^{-\lambda^* \tau_x}$, $|x| = n$. For better transparency, let us normalize these values to get a probability distribution on the n -th level of the tree: $p_x := \frac{T_x}{\sum_{|z|=n} T_z}$ (for $|x| = n$).
2. Also generate the random variables Θ_x for $|x| = n$, which are independent of the p_x .
3. Now y_n is chosen from the points $|x| = n$ according to the distribution μ_n , so the weight given to some x is

$$\frac{\Delta_x}{\sum_{|z|=n} \Delta_z} = \frac{T_x \Theta_x}{\sum_{|z|=n} T_z \Theta_z} = \frac{p_x \Theta_x}{\sum_{|z|=n} p_z \Theta_z}.$$

So, having the values p_x fixed, the value $X_n = \Theta_{y_n}$ is the result of a size-biased sampling from the independent random variables Θ_x , $|x| = n$, with additional weights p_x – just like in the context of Lemma 4.2 and Lemma 4.3.

Now we can prove (20). Let ϕ be a fixed bounded continuous function on \mathbb{R}^+ , let M_ϕ be an upper bound of $|\phi|$, and let $m_\phi = \int_{\mathbb{R}^+} \phi d\nu$ (which satisfies $|m_\phi| \leq M_\phi$). Let $\varepsilon > 0$ be arbitrary.

Choose $\delta > 0$ according to Lemma 4.3 so that if all the p_x on some level $|x| = n$ are at most δ , then

$$|\mathbf{E}(\phi(X_n) \mid \{p_x\}) - m_\phi| < \varepsilon.$$

Lemma 4.4 implies that there exists an n_0 such that for all $n > n_0$,

$$\mathbf{P}(\max\{p_x : |x| = n\} > \delta) < \frac{\varepsilon}{2M_\phi}.$$

Let $\Omega_{n,\delta}$ denote the event that $\max\{p_x : |x| = n\} \leq \delta$. For $n > n_0$ we get

$$\begin{aligned} |\mathbf{E}(\phi(X_n)) - m_\phi| &\leq \int |\mathbf{E}(\phi(X_n) - m_\phi \mid \{p_x\})| d\mathbf{P} = \\ &= \int_{\Omega_{n,\delta}^c} |\mathbf{E}(\phi(X_n) - m_\phi \mid \{p_x\})| d\mathbf{P} + \int_{\Omega_{n,\delta}} |\mathbf{E}(\phi(X_n) - m_\phi \mid \{p_x\})| d\mathbf{P} \leq \\ &\leq 2M_\phi \mathbf{P}(\Omega_{n,\delta}^c) + \int_{\Omega_{n,\delta}} \varepsilon d\mathbf{P} \leq \varepsilon + \varepsilon = 2\varepsilon. \end{aligned}$$

□

4.3 Weak Continuity of the Transition Kernel

This section is devoted to the proof of Lemma 3.15.

Proof of Lemma 3.15. We first show in Lemma 4.5 that the transition kernel P can be written as $(\eta P)(B) = \int_{\mathbb{R}^+} \int_B k(t, s) ds d\eta(t)$ where the kernel function $k(t, s)$ is continuous in the first variable (actually it is continuous in both variables). Lemma 4.6 – which is a pure probability statement – says that such a kernel is continuous with respect to weak convergence of measures. □

In the lemma, we show a little more than what is needed for the above proof. In particular, we also show that the kernel function $k(t, s)$ is nowhere zero on $\mathbb{R}^+ \times \mathbb{R}^+$, because this is used in the proof of Proposition 3.13.

Lemma 4.5. *The transition kernel P can be written as $(\eta P)(B) = \int_{\mathbb{R}^+} \int_B k(t, s) ds dt$ where the kernel function $k(t, s)$ is continuous in both variables (in its domain $(t, s) \in \mathbb{R}^+ \times \mathbb{R}^+$), and strictly positive.*

Proof. For the time of the proof, let Θ and Θ' denote two consecutive values of the process, say $\Theta := X_n = \Theta_{y_n}$, $\Theta' = X_{n+1} = \Theta_{y_{n+1}}$. So the kernel function $k(t, s)$ is just the conditional density of Θ' (as a function of s), under the condition $\Theta = t$. So

$$k(t, s) = \frac{\rho(t, s)}{\pi(t)},$$

where $\rho(t, s)$ is the joint density of the pair (Θ, Θ') , and $\pi(t)$ is its first marginal – that is, the density of Θ .

We know from Lemma 3.12 that Θ is indeed absolutely continuous w.r.t. Lebesgue measure, and the density π is continuous and nonzero on \mathbb{R}^+ . Knowing this, we now show that $\rho(t, s)$ is also continuous in both variables and nonzero on $\mathbb{R}^+ \times \mathbb{R}^+$, which completes the proof.

We restrict to the case $K = 2$. The case of a general $K < \infty$ causes no additional difficulty other than messy notation. Following the construction of the tree in Section 3.3, we start with $\sigma_1, \sigma_2, \Theta_1, \Theta_2$ independent, with σ_i being exponentially distributed with parameter $w(i-1)/\lambda^*$ and

Θ_i being distributed as Θ ($i = 1, 2$). We introduce the temporary notation $S_i = e^{-\lambda^* \sigma_i}$ and denote its density by g_i . Explicit calculation gives that

$$g_i(u) = \frac{w(i-1)}{\lambda^*} u^{\frac{w(i-1)}{\lambda^*}-1} \mathbb{1}_{(0,1)}(u), \quad (21)$$

of which we will only use that $u g_1(u)$ is bounded.

Denote the joint density of $(S_1, S_2, \Theta_1, \Theta_2)$ by

$$f(u_1, u_2, t_1, t_2) = g_1(u_1)g_2(u_2)\pi(t_1)\pi(t_2).$$

We define

$$\Theta = S_1 \Theta_1 + S_1 S_2 \Theta_2 = S_1(\Theta_1 + S_2 \Theta_2).$$

To get the appropriate joint distributions, in the probability vector $(S_1, S_2, \Theta_1, \Theta_2)$ we replace S_1 by Θ , so let us denote the joint density of $(\Theta, S_2, \Theta_1, \Theta_2)$ by \tilde{f} . The density transformation formula gives

$$\tilde{f}(t, u_2, t_1, t_2) = \frac{1}{t_1 + u_2 t_2} f\left(\frac{t}{t_1 + u_2 t_2}, u_2, t_1, t_2\right) = \frac{1}{t} \frac{t}{t_1 + u_2 t_2} g_1\left(\frac{t}{t_1 + u_2 t_2}\right) g_2(u_2) \pi(t_1) \pi(t_2).$$

According to the construction, Θ' is chosen to be either Θ_1 or Θ_2 , with conditional probabilities (given $(S_2, \Theta_1, \Theta_2)$ and conditionally independently of Θ)

$$\begin{aligned} \mathbf{P}(\Theta' = \Theta_1 | S_2, \Theta_1, \Theta_2) &= \frac{\Theta_1}{\Theta_1 + S_2 \Theta_2}, \\ \mathbf{P}(\Theta' = \Theta_2 | S_2, \Theta_1, \Theta_2) &= \frac{S_2 \Theta_2}{\Theta_1 + S_2 \Theta_2}. \end{aligned}$$

So the joint density of (Θ, Θ') is

$$\begin{aligned} \rho(t, s) &= \iint_{\mathbb{R}^2} \frac{t_1}{t_1 + u_2 t_2} \tilde{f}(t, u_2, s, t_2) dt_2 du_2 + \iint_{\mathbb{R}^2} \frac{u_2 t_2}{t_1 + u_2 t_2} \tilde{f}(t, u_2, t_1, s) dt_1 du_2 = \\ &= \iint_{\mathbb{R}^2} \bar{f}_1(t, s, u_2, t_2) dt_2 du_2 + \iint_{\mathbb{R}^2} \bar{f}_2(t, s, u_2, t_1) dt_1 du_2 \end{aligned} \quad (22)$$

All there is left is to show that both integrals on the right hand side are continuous and nonzero for $(t, s) \in \mathbb{R}^+ \times \mathbb{R}^+$. Now the integrands \bar{f}_1 and \bar{f}_2 are not exactly continuous, but they are continuous *on their supports*.³ On the other hand, for every $(t, s) \in \mathbb{R}^+ \times \mathbb{R}^+$ the support of each integrand is a nice set (described in the footnote) with a boundary of Lebesgue measure zero. That is, for *every* $(t_0, s_0) \in \mathbb{R}^+ \times \mathbb{R}^+$,

$$\bar{f}_1(t, s, u_2, t_2) \xrightarrow{(t,s) \rightarrow (t_0, s_0)} \bar{f}_1(t_0, s_0, u_2, t_2) \text{ for Lebesgue-a.e. } (u_2, t_2) \in \mathbb{R}^2.$$

To get the desired continuity of the first integral by the Lebesgue dominated convergence theorem, we only need to find an integrable (in (u_2, t_2)) uniform (in (t, s) near (t_0, s_0)) upper bound for

$$\bar{f}_1(t, s, u_2, t_2) = \frac{s}{s + u_2 t_2} \frac{1}{t} \frac{t}{s + u_2 t_2} g_1\left(\frac{t}{s + u_2 t_2}\right) g_2(u_2) \pi(s) \pi(t_2).$$

The first factor is at most 1, and the product $\frac{t}{s + u_2 t_2} g_1\left(\frac{t}{s + u_2 t_2}\right)$ is bounded because $u g_1(u)$ is bounded due to (21). So we have

$$\bar{f}_1(t, s, u_2, t_2) \leq C \frac{1}{t} \pi(s) g_2(u_2) \pi(t_2) \leq C \left(\frac{1}{t_0} + 1\right) (\pi(s) + 1) g_2(u_2) \pi(t_2)$$

if (t, s) is close enough to (t_0, s_0) , since $\frac{1}{t} \pi(s)$ is continuous in (t_0, s_0) . This upper bound is clearly integrable in (u_2, t_2) , so the dominated convergence theorem ensures that the integral is also continuous.

³The supports of the two integrands are actually not the same. Both of them are characterized by the system of inequalities $\{0 < t_1, t_2; 0 < u_2 < 1; 0 < \frac{t}{t_1 + u_2 t_2} < 1\}$, but with the choice $s = t_1$ or $s = t_2$, respectively.

The second integral in (22) can be shown to be continuous in exactly the same way. Thus the continuity of $k(t, s)$ is proven.

To get that $\rho(t, s)$ (and thus $k(t, s)$) is strictly positive on $\mathbb{R}^+ \times \mathbb{R}^+$, we only need to note that the support of the integrand is nonempty for every $(t, s) \in \mathbb{R}^+ \times \mathbb{R}^+$ in both integrals on the right hand side of (22). This comes again from (14), which shows that any pair of positive values is possible for (Θ, Θ_1) (in case of the first integrand) or for (Θ, Θ_2) (in case of the second integrand). (See the footnote 3 for explicit formulae.) The integrands are of course also non-negative, so both integrals are positive. \square

Lemma 4.6. *Let $k : \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow [0, \infty)$ be a function continuous in the first variable, such that for every $t \in \mathbb{R}^+$ the function $k(t, \cdot)$ is a probability density on \mathbb{R}^+ – that is, $\int_{\mathbb{R}^+} k(t, s) ds = 1$. Let the operator P be defined on Borel probability measures of \mathbb{R}^+ by*

$$(\eta P)(B) := \int_{\mathbb{R}^+} \int_B k(t, s) ds d\eta(t)$$

for every Borel probability measure η on \mathbb{R}^+ and every Borel set $B \subset \mathbb{R}^+$. Then P is continuous with respect to weak convergence of measures.

This lemma is an easy consequence of the following:

Lemma 4.7. *Let $k : \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow [0, \infty)$ be a function as in Lemma 4.6, and for every $t \in \mathbb{R}^+$ let K_t denote the measure on \mathbb{R}^+ with density $k(t, \cdot)$. Then if t_n is a sequence in \mathbb{R}^+ converging to t , then K_{t_n} converges to K_t weakly.*

Proof. By assumption, $\{k(t_n, \cdot)\}_{n=1}^\infty$ is a sequence of density functions converging pointwise to the density function $k(t, \cdot)$. This implies weak convergence of the corresponding measures through the Fatou lemma: for any Borel set $B \subset \mathbb{R}^+$

$$\liminf_{n \rightarrow \infty} K_{t_n}(B) = \liminf_{n \rightarrow \infty} \int_B k(t_n, s) ds \stackrel{\text{Fatou}}{\geq} \int_B \liminf_{n \rightarrow \infty} k(t_n, s) ds = \int_B k(t, s) ds = K_t(B),$$

similarly

$$\liminf_{n \rightarrow \infty} K_{t_n}(B^c) \geq K_t(B^c),$$

which implies

$$\limsup_{n \rightarrow \infty} K_{t_n}(B) = 1 - \liminf_{n \rightarrow \infty} K_{t_n}(B^c) \leq 1 - K_t(B^c) = K_t(B).$$

These together give

$$K_{t_n}(B) \rightarrow K_t(B).$$

\square

Proof of Lemma 4.6. Let $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}$ be bounded and continuous and let η_n be a sequence of measures on \mathbb{R}^+ converging weakly to η . By the definition of P ,

$$\begin{aligned} \int_{\mathbb{R}^+} \phi d(\eta_n P) &= \int_{\mathbb{R}^+ \times \mathbb{R}^+} k(t, s) \phi(s) d(\eta_n(t) \times \text{Leb}(s)) = \\ &= \int_{\mathbb{R}^+} \left[\int_{\mathbb{R}^+} k(t, s) \phi(s) ds \right] d\eta_n(t). \end{aligned}$$

The function

$$\bar{\phi}(t) := \int_{\mathbb{R}^+} k(t, s) \phi(s) ds$$

is obviously bounded, and also continuous: this is exactly the statement of Lemma 4.7. But then the weak convergence of η_n to η means exactly that

$$\int_{\mathbb{R}^+} \bar{\phi}(t) d\eta_n(t) \rightarrow \int_{\mathbb{R}^+} \bar{\phi}(t) d\eta(t),$$

so we have

$$\int_{\mathbb{R}^+} \phi d(\eta_n P) \rightarrow \int_{\mathbb{R}^+} \bar{\phi}(t) d\eta(t) = \int_{\mathbb{R}^+} \phi d(\eta P)$$

for every bounded continuous ϕ , which is exactly what we want to prove. \square

5 Computation of the Entropy

Proof of Theorem 2.5. We know that $\frac{1}{n}H_n = -\frac{1}{n} \sum_{|x|=n} \Delta_x \log \Delta_x$ converges almost surely to some constant h , and this constant is equal to the limit of the expected values. For this section we use the shorthand notation already introduced in (17),

$$T_x = e^{-\lambda^* \tau_x}. \quad (23)$$

To compute h , first observe that

$$\begin{aligned} \mathbf{E} \sum_{|x|=n} \Delta_x \Theta \log(\Delta_x \Theta) &= \mathbf{E} \left(\sum_{|x|=n} \Theta \Delta_x \log \Delta_x \right) + \mathbf{E} \left((\Theta \log \Theta) \sum_{|x|=n} \Delta_x \right) = \\ &= \mathbf{E} \left(\Theta \sum_{|x|=n} \Delta_x \log \Delta_x \right) + \mathbf{E}(\Theta \log \Theta), \end{aligned}$$

where we have used that $\sum_{|x|=n} \Delta_x = 1$ by definition.

Next we observe that on the other hand, the same expression can be written as

$$\begin{aligned} \mathbf{E} \sum_{|x|=n} \Delta_x \Theta \log(\Delta_x \Theta) &= \mathbf{E} \sum_{|x|=n} T_x \Theta_x \log(T_x \Theta_x) = \\ &= \mathbf{E} \left(\sum_{|x|=n} \Theta_x T_x \log(T_x) \right) + \mathbf{E} \left(\sum_{|x|=n} T_x \Theta_x \log \Theta_x \right) = \\ &= \sum_{|x|=n} (\mathbf{E} \Theta_x) \mathbf{E}(T_x \log T_x) + \sum_{|x|=n} \mathbf{E}(T_x) \mathbf{E}(\Theta_x \log \Theta_x) = \\ &= (\mathbf{E} \Theta) \mathbf{E} \sum_{|x|=n} (T_x \log T_x) + \mathbf{E}(\Theta \log \Theta) \mathbf{E} \left(\sum_{|x|=n} T_x \right), \end{aligned}$$

where we have used that for any $x \in \mathcal{N}$, Θ_x and τ_x are independent. Recall that $\mathbf{E} \left(\sum_{|x|=n} T_x \right) = 1$.

Since (4) implies that $\mathbf{E}(\Theta \log \Theta) < \infty$, comparing the two formulae gives the conclusion

$$\mathbf{E} \left(\Theta \sum_{|x|=n} \Delta_x \log \Delta_x \right) = (\mathbf{E} \Theta) \mathbf{E} \left(\sum_{|x|=n} T_x \log T_x \right). \quad (24)$$

We compute the right-hand side with an induction,

$$\begin{aligned} A_n &:= \mathbf{E} \left(\sum_{|x|=n} T_x \log T_x \right) = \mathbf{E} \left(\sum_{|y|=n-1} \sum_{i=1}^K T_{yi} \log T_{yi} \right) = \\ &= \left(\mathbf{E} \sum_{i=1}^K e^{-\lambda^*(\tau_{yi}-\tau_y)} \right) \mathbf{E} \left(\sum_{|y|=n-1} T_y \log T_y \right) + \\ &= \left(\mathbf{E} \sum_{|y|=n-1} T_y \right) \mathbf{E} \left(\sum_{i=1}^K e^{-\lambda^*(\tau_{yi}-\tau_y)} \log e^{-\lambda^*(\tau_{yi}-\tau_y)} \right) = \\ &= A_{n-1} + \mathbf{E} \left(\sum_{i=1}^K T_i \log T_i \right), \end{aligned}$$

so

$$A_n = n\mathbf{E}\left(\sum_{i=1}^K T_i \log T_i\right).$$

Now write this back to (24) to get

$$\mathbf{E}\left(\Theta \frac{1}{n} H_n\right) = (\mathbf{E}\Theta) \mathbf{E}\left(-\sum_{i=1}^K T_i \log T_i\right).$$

Since $\lim \frac{1}{n} H_n = h$ almost surely and $\mathbf{E}\Theta < \infty$, we can apply the dominated convergence theorem if we check that $\frac{1}{n} H_n$ is bounded. This follows from the standard upper bound for entropy of measures on the finite set $\{x \in \partial\mathcal{N} : |x| = n\}$, which has K^n elements, coming from the Jensen inequality:

$$\begin{aligned} H_n &= -\sum_{|x|=n} \mu_n(\{x\}) \log \mu_n(\{x\}) = \int_{\{x \in \partial\mathcal{N} : |x|=n\}} \log \frac{1}{\mu_n(\{x\})} d\mu_n(x) \stackrel{\text{Jensen}}{\leq} \\ &\leq \log \int_{\{x \in \partial\mathcal{N} : |x|=n\}} \frac{1}{\mu_n(\{x\})} = \log \sum_{|x|=n} \mu_n(\{x\}) \frac{1}{\mu_n(\{x\})} = \log K^n = n \log K, \end{aligned}$$

so $\frac{1}{n} H_n \leq \log K$. Now dominated convergence gives

$$h = \mathbf{E}\left(-\sum_{i=1}^K T_i \log T_i\right).$$

Recalling (23), the proof of the theorem is complete. \square

Remark 5.1. *This value can be explicitly calculated, as soon as the weight function is given, since the τ_i variables are the sum of independent, exponentially distributed random variables with parameters $(w(j))_{j=0}^{i-1}$. Alternatively, with the function \widehat{g} defined in (2),*

$$h = \lambda^* \frac{d\widehat{g}(\lambda)}{d\lambda} \Big|_{\lambda=\lambda^*}.$$

6 Outlook

The present result is restricted to the $K < \infty$ case, i.e. when a vertex can only have finitely many children. This property is used in three places. First, Theorem 2.3 relies on Lemma 3.18, which is a very rough estimate working for finite K only. Second, in the proof of Theorem 2.5 we use the fact that $\frac{1}{n} H_n$ is bounded – which is also certainly false for $K = \infty$. Third, showing the continuity of the density π and the transition kernel function k (in lemmas 3.12 and 4.5) is easier using the fact that the sum in (5) is finite. With more care, these could possibly be generalized for the $K = \infty$ case, so the main result about the Hausdorff dimension, Theorem 2.4 could be shown in greater generality. However, not having the explicit formula of Theorem 2.5 is a serious drawback. We believe that the problem can be solved – and the validity of the explicit formula can be shown – for a large class of rate functions with $K = \infty$ by a detailed analysis of the transition kernel P . Such an analysis could be avoided in the present paper by the study of the limiting distribution in Section 4.2. We plan to return to that in the future.

Acknowledgements

We gratefully thank Balázs Ráth for the simple proof of Lemma 4.2. We are also grateful to an anonymous referee for many useful suggestions that helped improve the quality of the paper. A. Rudas acknowledges the support of OTKA grant K60708. I. P. Tóth acknowledges the support of OTKA grants PD73609 and K71693, and is also grateful to the European Research Council for support.

References

- [1] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [2] J. Berestycki. Multifractal spectra of fragmentation processes. *Journal of Statistical Physics*, 113(3):411–430, 2003.
- [3] J. Bertoin. *Random fragmentation and coagulation processes*. Cambridge Univ Pr, 2006.
- [4] J. D. Biggins. Martingale convergence in the branching random walk. *Journal of Applied Probability*, 14(1):pp. 25–37, 1977.
- [5] Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. The degree sequence of a scale-free random graph process. *Random Structures Algorithms*, 18(3):279–290, 2001.
- [6] Joseph Leo Doob. *Stochastic Processes*. Wiley, 1953.
- [7] T. Duquesne. Packing and Hausdorff measures of stable trees. *Lévy Matters I*, pages 93–136, 2010.
- [8] Thomas Duquesne and Jean-Francois Le Gall. Probabilistic and fractal aspects of Lévy trees. *Probability Theory and Related Fields*, 131:553–603, 2005.
- [9] Kenneth Falconer. *Techniques in Fractal Geometry*. Wiley, 1997.
- [10] B. Haas and G. Miermont. The genealogy of self-similar fragmentations with negative index as a continuum random tree. *Electronic Journal of Probability*, 9(paper 4):57, 2004.
- [11] B. Haas and G. Miermont. Scaling limits of Markov branching trees, with applications to Galton-Watson and random unordered trees. *Arxiv preprint arXiv:1003.3632*, 2010.
- [12] B. Haas, G. Miermont, J. Pitman, and M. Winkel. Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models. *The Annals of Probability*, 36(5):1790–1837, 2008.
- [13] Peter Jagers. *Branching processes with biological applications*. Wiley-Interscience [John Wiley & Sons], London, 1975. Wiley Series in Probability and Mathematical Statistics —Applied Probability and Statistics.
- [14] P. L. Krapivsky and S. Redner. Organization of growing random networks. *Phys. Rev. E*, 63(6):066123, May 2001.
- [15] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Phys. Rev. Lett.*, 85(21):4629–4632, Nov 2000.
- [16] R. Lyons. A simple path to Biggins’ martingale convergence for branching random walk. In K.B. Athreya and P. Jagers, editors, *Classical and modern branching processes*, The IMA volumes in mathematics and its applications. Springer, 1997.
- [17] R. Lyons, R. Pemantle, and Y. Peres. Conceptual proofs of $l \log l$ criteria for mean behavior of branching processes. *The Annals of Probability*, 23(3):1125–1138, 1995.
- [18] T. F. Móri. On random trees. *Studia Sci. Math. Hungar.*, 39(1-2):143–155, 2002.
- [19] Roberto Oliveira and Joel Spencer. Connectivity transitions in networks with super-linear preferential attachment. *Internet Math.*, 2(2):121–163, 2005.
- [20] Anna Rudas and Bálint Tóth. Random tree growth with branching processes - a survey. In B Bollobás, R Kozma, and D Miklós, editors, *Handbook of Large-Scale Random Networks*, volume 18 of *Bolyai Society Mathematical Studies*, chapter 4. Springer, 2007.
- [21] Anna Rudas, Bálint Tóth, and Benedek Valkó. Random trees and general branching processes. *Random Struct. Algorithms*, 31(2):186–202, 2007.